Proceedings of the Annual General Donald R. Keith Memorial Conference
West Point, New York, USA
April 28, 2022
A Regional Conference of the Society for Industrial and Systems Engineering

Utilizing Machine Learning Models to Predict Success in Special Operations Assessment

Anna Vinnedge, Blake Schwartz, Daniel Baller, and Elise Dykhuis

Department of Mathematical Sciences
United States Military Academy
West Point, NY 10996

Corresponding author's Email: anna.vinnedge@westpoint.edu

Author Note: Sincerest gratitude to the 75th Ranger Regiment and their Regimental Department of Operational Psychology, for funding this research and providing necessary resources and data. Additional thanks to CPT Deverill, for assistance in data collection and cleaning.

Abstract: The 75th Ranger Regiment is an elite unit within the United States Special Operations Command tasked with some of the most physically and mentally challenging missions. Entry to the unit is based on an assessment process called Ranger Assessment and Selection Program (RASP), which consists of a variety of tests and challenges of strength, intellect, and grit. This study explores the psychological and physical profiles of candidates who attempt to pass RASP. Using a Random Forest Artificial Intelligence model, and a penalized logistic regression model, we identify initial entry characteristics that are predictive of success in RASP. We focus on the differences between racial sub-groups and military occupational specialties (MOS) to provide information for recruiters to identify individuals in underrepresented groups who are likely to succeed into the selection process. Several specific variables are de-identified in this paper to maintain equity in future candidate preparation, but the results have been provided to the Ranger Regiment to use as an informative tool.

Keywords: Random Forest, Logistic Regression, Special Operations, Ranger Regiment, Machine Learning

1. Introduction

Both machine learning models and logistic regression models have been widely used in predicting success in varying types of employment (Couronne, 2018). In this study, we harness the capabilities of both modeling methods to identify factors contributing to the likelihood of success in the assessment and selection process for the 75th Ranger Regiment The 75th Ranger Regiment is an elite unit within the United States Special Operations Command, conducting a range of special operations to achieve mission success. In order to become a member of the 75th Ranger Regiment, every candidate must pass the Ranger Assessment and Selection Program (RASP). This study examines traits that are predictive of success in RASP. This is a physically and mentally demanding process which, on average, has a pass rate of approximately 35% (A. Brickhouse, personal communication, June 17, 2021). Utilizing predictive modeling techniques, we look for strategies to increase the rate of those who pass by identifying who will be more likely to succeed before they even start. This information can lead to an increase in the pass rate of RASP benefit the Ranger Regiment because less time and resources would be used it assessing candidates who ultimately do not pass.

1.1 Literature Review

Other researchers have successfully used logistic regression and machine learning for similar purposes in a variety of applications. In one study conducted in Spain, a machine learning model was used to test predictability of success in employment of university graduates upon obtaining their bachelor's degree. The data included graduates from a variety of schools and various employers. Statistics on the subjects were collected using a survey, mainly focusing on demographics, educational competencies, and academic backgrounds. After data cleaning and normalization. A random forest algorithm was selected to produce a binary classification result. The model was trained with 33.3% of the dataset, and then used to obtain the most important features. The final model produced 71% precision. The researchers also conducted a cluster analysis to remove data that was highly correlated to account for potential overfitting. We use cluster analysis in our data cleaning process for similar purposes (Francisco, 2018).

Another study used logistic regression, decision trees, and random forest models to construct early warning models of student success in Physics 1 and 2 during their undergraduate education. Variables such as homework grades, GPA, and demographics allowed for the models to predict if a student would receive less than a B in physics 1 with 73% accuracy and 81% accuracy for physics 2. Further findings showed that institutional variables, such as cumulative GPA were the most significant predictors in early performance while homework grades became more predictive as the semester progressed. Surprisingly, demographic factors such as gender and race did not show significant importance in predictions in any of the models (Zabriskie, 2019).

There has also been extensive research done comparing and contrasting the utility of different types of predictive models, including random forest and logistic regression, which are both used in this study. One large benchmark study, which looked at the prediction success on 243 real datasets, compared random forest and logistic regression on a binary predictor, similar to our pass-fail variable. This study looked mainly at clinical trial data and concluded that random forest performed better than logistic regression in 69% of datasets in accuracy. Accuracy was measured using AUC (area under curve) which measures probability of correct classification. The Brier score was also used which measures the deviation between true classification and predicted probability. Results showed that random forest performed better under the AUC measure 72.3% of the time and 71.5% for the Brier. The study also emphasizes the importance of inclusion criteria and size of datasets (Couronne, 2018).

2. Methods

2.1 Data Analysis

Prior to developing predictive models, it was necessary to collect, interpret, and organize any data that could be relevant to this research.

2.1.1 Data Collection

The dataset used in this analysis was collected from January of 2017- June of 2021 on all candidates who attempted RASP. The initial dataset contained approximately 11,000 entries of 370 variables. 40% of recorded variables came from scores from psychometric testing. Prior to beginning RASP, candidates take a series of psychological tests which rank different character traits and tendencies amongst the population of American people of the same age and sex. In addition to psychometric data, demographic data, physical fitness scores, and additional variables were also recorded. The data was provided by the head data analyst for RASP as a data frame saved in Microsoft access. Visiting the Ranger Regiment, we conducted individual interviews on candidates who passed RASP which provided us with a holistic view of the experience of individuals and personal motivators of success.

2.1.2 Data Cleaning

The initial dataset contained a significant amount of missing data. The first step in the data cleaning process was removing all entries with missing values within the psychometric data because our primary concern was looking at predictive psychometric factors. This removed half of the individuals from our data set. Next, we explored the scheduling of RASP, identifying which data was recorded prior to the start of RASP. Any variables (such as fitness assessments which took place during the training) were dropped from the data since our goal was to predict success prior to starting RASP. For demographic variables with high levels of N/A's, we looked to see if default entries were plausible. For example, in the entry for languages spoken, we chose to default N/A's to "English" as being the only language spoken. This was done for 8 categorical variables and 4 numeric variables. Additionally, we grouped some levels of categorical variables into broader categories for the purpose of preventing potential overfitting. For example, we transformed the variable state of origin into five regions of the United States. As a final step, we removed any observations with missing values in the variables selected for modeling.

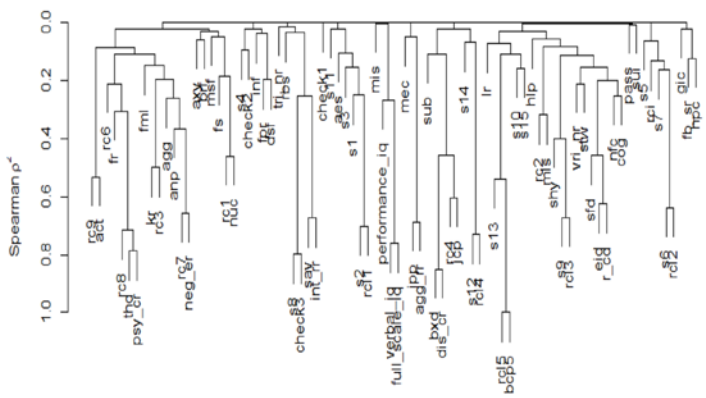

Figure 1. Correlation tree of psychometric testing variables. Trees such as this one were used to identify closely correlated variables.

Once the data was cleaned, we conducted a correlation analysis of the psychometric data. The intent was to identify if any psychological factors correlated to such an extent that including them would be unnecessary and may contribute to overfitting (refer to Figure. 1). Using correlation coefficients, we decided to drop one of the two metrics in any instance of the coefficient exceeding 0.8, which is indicative of a very high correlation. By the end of the data cleaning process, we were left with 3,482 observations of 111 variables (refer to Figure. 2).

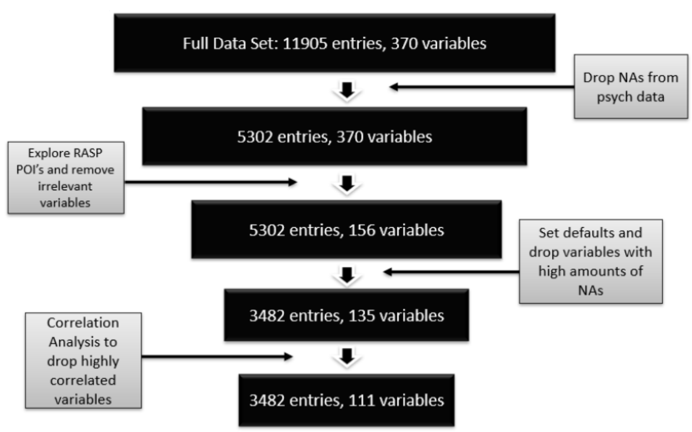

Figure 2. Flowchart of data cleaning process.

2.2 Methods of Analysis

2.2.1 Random Forest Model

Random forest models are a way of averaging multiple decision trees, trained on different subsections of the same training set, with the goal of reducing the problem of over-fitting from an individual decision tree. Random forests are a machine learning method for classification that outputs the mode outcome (pass or fail in the case of our model) from the collection of individual trees (Yiu, 2019).

Random Forest is one of the most widely used machine learning algorithms for classification. It can also be used for regression models (i.e. continuous target variable) but it mainly performs well on classification models (i.e. categorical target variable) (Ibid).

2.2.2 Penalized Logistic Regression

Logistic regression is a type of statistical analysis is often used for predictive analytics and modeling and extends to applications in machine learning. In this analytics approach, the dependent variable is finite or categorical: either A or B (binary regression) or a range of finite options A, B, C or D (multinomial regression) (Logistic).

Standard logistic regression models the probability of success by assigning coefficients to predictive variables. Penalized logistic regression imposes a penalty to the logistic model for having too many variables, shrinking coefficients of less contributive variables to zero (Kassambara, 2018).

3. Results

3.1 Comparison of models

Using RStudio and the TidyModels package, we implemented both a penalized logistic regression model and a random forest model, both of which used all 111 variables to make a binary prediction of passing or failing RASP. To measure success of the models, we examined accuracy and ROC-AUC (receiver operator characteristic-area under curve) metrics. By splitting the data into training and testing data, values from a confusion matrix can be used to calculate model performance in terms of true negatives (TN), true positives (TP), false negatives (FN) and false positives (FP) on our testing set (Ragan, 2018).

Accuracy = (TN + TP)/(TN+TP+FN+FP) = (Number of correct assessments)/(Number of all assessments) (1)

ROC-AUC is a performance measurement for classification problems at various threshold settings. ROC is a probability curve and AUC represents the ability of the model to distinguish between classes. It tells how much the model is capable of distinguishing between classes. A higher AUC measurement indicates that the model is better at predicting true positives and true negatives. We measured this value numerically and also visually with the two axis' being sensitivity and 1-specificity (Narkhede, 2018).

Sensitivity = TP/(TP + FN) = (Number of true positive assessment)/(Number of actual positive values) (2)

Specificity = TN/(TN + FP) = (Number of true negative assessment)/(Number of actual negative values) (3)

The calculations indicated that the random forest model was more successful at accurately predicting the correct binary outcome. Both models had high accuracy levels (.79 for penalized logistic and .80 for random forest) as well as promising ROC-AUC measures (.86 for penalized logistic and .88 for random forest)

Table 1. Model Success Measures for Penalized Logistic Regression Model

Metric	Value
Accuracy	0.7944891
ROC_AUC	0.8617775

Table 2. Model Success Measures for Random Forest Model

Metric	Value
Accuracy	0.7990815
ROC_AUC	0.8878728

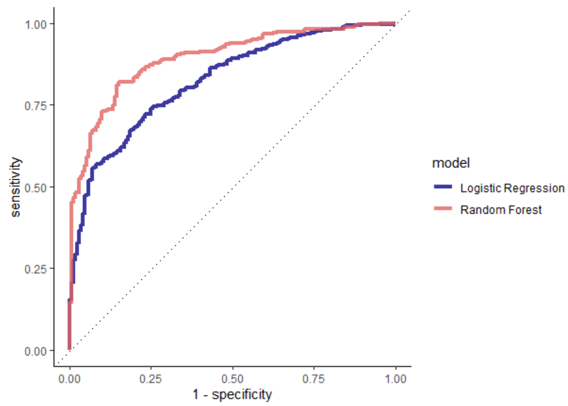

Figure 3. ROC-AUC Curves Comparing Penalized Logistic Regression Model and Random Forest Model

3.1.1 Analysis of Best Model

After comparing both models, we conclude that random forest has the highest likelihood of accurately predicting success in RASP. To provide insight into which variables are the most influential to success in RASP, we used the R VIP (variable importance plots) package to measure importance levels. This works with binary decision trees by, at each node, a single predictor is used to partition the data into two homogeneous groups. The chosen predictor is the one that maximizes some measure of improvement. The relative importance of predictor is the sum of the squared improvements over all internal nodes of the tree for which that predictor was chosen as the partitioning variable (Greenwell, 2020). Using this tool, we ranked the most important factors in predicting success in RASP.

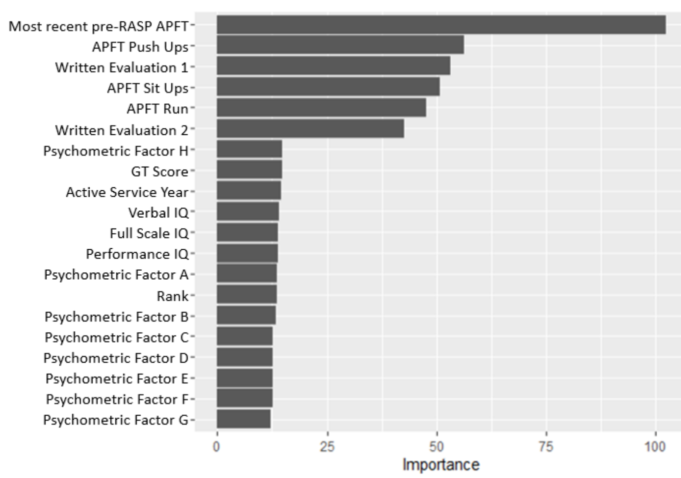

Figure 4. Rankings of most important predictive factors for success in RASP

After observing the rankings for the full model, we subset the data by race and MOS (military occupational specialty) to explore the differences in predictive factors between groups of specific interest to the Ranger Regiment. Table 3 shows the breakdown by race, with variables listed from most to least important (top 20 variables). Race was subset into white (n=3132), and other, referred to as “minority” (n=350). Table 4 shows the breakdown by MOS into four categories: high (n=2106), middle (n=200), low (n=917), and special (n=259). High includes all infantry MOS meaning those who specialize in close range land combat. Middle includes all other combat arms MOS such as field artillery, engineers, armor, or air defense artillery. Special refers to any special forces MOS, and low includes all non-combat (primarily logistic or combat support) MOS.

ISBN: 97819384962-2-6

5. Conclusions

5.1 Analysis of Results

The model for the entire data set shows that physical fitness holds significant weight as a predictor of success in RASP. IQ and additional written evaluations also show to be very important as an indicator of success. Several psychometric factors appear in the most important variables, specifics of which have been provided to the Ranger Regiment to utilize as they see useful.

Looking into the differences between racial minorities and white candidates, the most general difference is that some of the psychometric factors are more prevalent as important factors in racial minorities. It is difficult to draw general conclusions from the breakdown of variables by MOS, however, understanding which factors are important may be beneficial in recruiting underrepresented groups.

5.2 Implications and Future Steps

While this model is not necessarily meant as a tool to eliminate candidates who may be at risk of failure, it is still useful to be able to identify candidates with a higher likelihood of success for training and recruiting purposes. Additionally, it could help identify candidates who may need to go through additional screening or preparation prior to RASP to increase their success probability. Providing model results to candidates may also increase self-awareness of strengths and weaknesses as they decide if they would like to attempt RASP. This could increase the volume of capable candidates attending RASP which would provide economic benefit to the Ranger Regiment because they would not be spending money on training people who do not end up in the unit.

Most importantly, the results of this study can be used as a recruiting tool outside of the Ranger Regiment by providing information to recruiters for traits to look for when recommending soldiers for RASP. More analysis is necessary to know for certain, but it is possible that administering the psychological profiling and fitness testing to candidates before they report to the Ranger Regiment, or even to the Army as a whole, could help to identify promising candidates from around the Army. This is especially important in recruiting of low density MOS's which are in high demand and have a low pass rate for RASP. Similarly, as specialized units look to become more representative of the Army as a whole, these models can be used to identify excellent racial minority and female candidates, both highly underrepresented groups, who are likely to succeed in RASP.

6. References

Couronné, R., Probst, P., & Boulesteix, A. L. (2018). Random forest versus logistic regression: a large-scale benchmark experiment. *BMC bioinformatics*, 19(1), 1-14.

García-Peñalvo, F. J., Cruz-Benito, J., Martín-González, M., Vázquez-Ingelmo, A., Sánchez-Prieto, J. C., & Therón, R. (2018). Proposing a machine learning approach to analyze and predict employment and its factors.

Greenwell, B. M., & Boehmke, B. C. (2019). Variable Importance Plots: An Introduction to Vip.

Hugo, L. S. (2019, May 1). Predicting Employment Through Machine Learning. *National Association of Colleges and Employers*. https://www.naceweb.org/career-development/trends-and-predictions/predicting-employment-through-machine-learning/

Kassambara, A. (2018, March 11). *Penalized Logistic Regression Essentials in R: Ridge, Lasso and Elastic Net*. Statistical Tools for High-Throughput Data Analysis. http://www.sthda.com/english/articles/36-classification-methods-essentials/149-penalized-logistic-regression-essentials-in-r-ridge-lasso-and-elastic-net

Kirasich, K., Smith, T., & Sadler, B. (2018). Random forest vs logistic regression: binary classification for heterogeneous datasets. *SMU Data Science Review*, 1(3), 9.

Narkhede, S. (2022, March 5). *Understanding AUC - ROC Curve - Towards Data Science*. Medium. https://towardsdatascience.com/understanding-auc-roc-curve-68b2303cc9c5

Ragan, A. (2018, October 12). *Taking the Confusion Out of Confusion Matrices - Towards Data Science*. Medium. https://towardsdatascience.com/taking-the-confusion-out-of-confusion-matrices-c1ce054b3d3e

What is Logistic regression? | IBM. (n.d.). IBM. https://www.ibm.com/topics/logistic-regression

Zabriskie, C., Yang, J., DeVore, S., & Stewart, J. (2019). Using machine learning to predict physics course outcomes. *Physical Review Physics Education Research*, 15(2), 020120.